

平成 26 年度 地球シミュレータ利用報告 研究成果概要

1. 課題名

地球環境変動・保全に係わる全地球レベルでの微生物群集構造把握のためのゲノム情報基盤整備

A Large-scale Self-Organizing Map for Metagenome Studies for Surveillance of Microbial Community Structures

2. 課題責任者

池村淑道(長浜バイオ大学)

Toshimichi Ikemura

3. 課題の目的

地球環境の変動が生物生態系に多大な影響を与えると共に、生物生態系側も地球環境へ多大な影響を与えて来た。海洋に生息する植物プランクトン・光合成微生物類の様に、炭素循環に重要な役割を果たしているながらも、肉眼的な観察が困難な場合、関与している微生物類の生態系の実態、特に地球レベルでの生態については未知に残されて来た。更に重要な点は、光合成微生物類を例に取っても、地球環境への影響を知るには、個々の微生物種ではなく、微生物群集構造の把握が必須である。しかしながら、自然環境の微生物類の 99%以上が実験室では培養が困難であり、関与する微生物類の特定を困難にして来た。最近の DNA 配列の解読技術の飛躍的な高速化は、「メタゲノム解析法」と呼ばれる革新的な手法を確立させた。海洋を代表例として全地球レベルでの大規模なメタゲノム解析が進行し、超大量なゲノム配列が集積している。我々が開発して来た、一括学習型自己組織化マップ法(BLSOM)は、断片ゲノム配列を生物系統ごとに高精度に分離(自己組織化)する能力を持つ。解読された全ゲノム配列を対象に ES で大規模 BLSOM を作成・更新して行けば、メタゲノム解析で得られる大量配列をマップすることで、極限環境や海水等の様々な環境中で生息する生物群集の全体像の把握が可能になり、加えて、環境保全・浄化に役立つ新規微生物類やそれらの保持する有用遺伝子類を発掘できる。

4. 今年度当初の研究計画

これまでの研究開発で、大規模ゲノム配列解読に対応できる BLSOM システムを確立し、大学を含む公的な研究機関からの環境ゲノム配列の系統推定解析の依頼に応じてきた。多様な環境試料を対象にしたメタゲノム解析は、公的研究機関のみならず、産業界からも注目を集めている。今年度時点で DNA データベースに収録されている原核生物、真核生物、ウイルス等の全ゲノム配列データを用いた大規模 BLSOM を、年間で少なくとも一回は更新を行う。さらに、より大量なゲノム配列データに対応するために、より効率的なアルゴリズムの開発を行う。

5. 研究計画に沿った利用状況

環境中で生育する新規性の高い、未知の微生物由来のメタゲノム配列の生物系統の推定を行うには、現時点で研究の進んだ全ての既知の生物種由来のゲノム配列で大規模 BLSOM を作成しておく必要がある。大規模 BLSOM の更新を行い、この BLSOM を用いて大量メタゲノム配列の系統推定を行うためのソフトウェア PEMS の公開を行い、ES での研究成果の普及を図っている。さらに、地球環境の把握のため、極域や深海に生息する微生物に着目したゲノム・メタゲノム配列データを対象に、BLSOM 解析を実施している。

6. 今年度得られた成果、および達成度

<成果>

1) 全ゲノム配列データを対象にした大規模 BLSOM の更新とその利用

環境中で生育する新規性の高い、未知の微生物由来のメタゲノム配列の生物系統を推定する目的で、本年度 9 月の段階で公開されていた全微生物のゲノム配列データを対象に、大規模 BLSOM を作成した。海洋由来試料のメタゲノム研究として、アカシオ発生の原因やそれを防ぐ方策を研究するための基盤情報を得る目的で、作成した大規模 BLSOM を用いて微生物叢の解明を行った。我が国の漁業や環境保全に重要な役割を果たす研究のプロトタイプ確立と考えられる。国立極地研究所と国立遺伝学研究所の研究グループとの共同研究として、北極・南極に生息する微生物ゲノム解読プロジェクトにおいて、ES で作成した大規模 BLSOM を活用し、寒冷地に適応する際に獲得した遺伝子の検出とその機能推定を実施した。

海洋研究開発機構が先端的な研究を進めている深海底を含む極限環境における難培養性古細菌ゲノムについても、その古細菌の生育環境を予測するために、ES で作成した大規模 BLSOM を活用した共同研究を実施した。

2) 大規模ゲノム解析のための自己圧縮 BLSOM の開発

次世代シーケンサの登場以降、様々な生物のゲノム配列データの増加は目を見張るほどである。そのため、大量の配列データを包括的かつ俯瞰的に分析できる新たな手法が必要とされる。我々が開発してきた、連続塩基組成のみに着目してゲノム配列断片を高精度にクラスタリングし可視化する一括学習型自己組織化マップ (BLSOM) は、数十億塩基に及ぶような大量なデータを解析することができる。しかしながら、最近に急増しているゲノム配列データを対象にした大規模 BLSOM を作成するには、地球シミュレータを用いても多大な計算資源と時間を要する。そこで、この大規模解析を高速に行うための自己圧縮 BLSOM を開発した。自己圧縮 BLSOM の原理は、入力データの分類に基づいた BLSOM を階層的に構築することである。はじめに入力データをデータの分類によって分割し、それぞれ 1 階層目の BLSOM で特徴抽出してデータ数を減らす。そして 1 階層目の BLSOM で得られた特徴データを元に 2 階層目の BLSOM を行う。我々は、原核生物の完全長ゲノムを用いて SC-BLSOM と BLSOM の比較を行った。その結果、SC-BLSOM は大幅に計算時間を短縮することができ、クラスタリング能力も BLSOM より強力であった。SC-BLSOM は大量のゲノム配列データから効率よく知識を得るためのより相応しい手法であり、ES を活用することで、加速度的に増加し続ける全ゲノム配列データに対する大規模 BLSOM 解析をより効率的に行うことができる。

<達成度>

(年度当初の研究計画を全て達成した場合を 100% / 複数の目標があった場合は、それぞれについて達成度を数値で記載)

BLSOM マップを用いてメタゲノム配列に対する系統推定に利用するための全ゲノム配列を対象にした BLSOM マップの更新は予定通り実施することができ、達成度は当初予定の 80%程度であった。

7. 計算機資源の利用状況

<計算機資源の利用状況>

(計画的に計算機資源を利用できているか、状況を記載)

BLSOM マップを用いてメタゲノム配列に対する系統推定に利用するための全ゲノム配列を対象にした BLSOM マップの更新は予定通り実施することができたが、より高速化を目指した大規模ゲノム解析のための自己圧縮 BLSOM のアルゴリズム、ならびに、条件検討に時間がかかり、自己圧縮 BLSOM の大規模解析を行うことができなかった。

<チューニングによる成果>

(ベクトル化、並列化チューニング等、計算機資源を有効利用するために行ったこととその効果を記載)

これまでに、ES2 上でのチューニングを終えている。

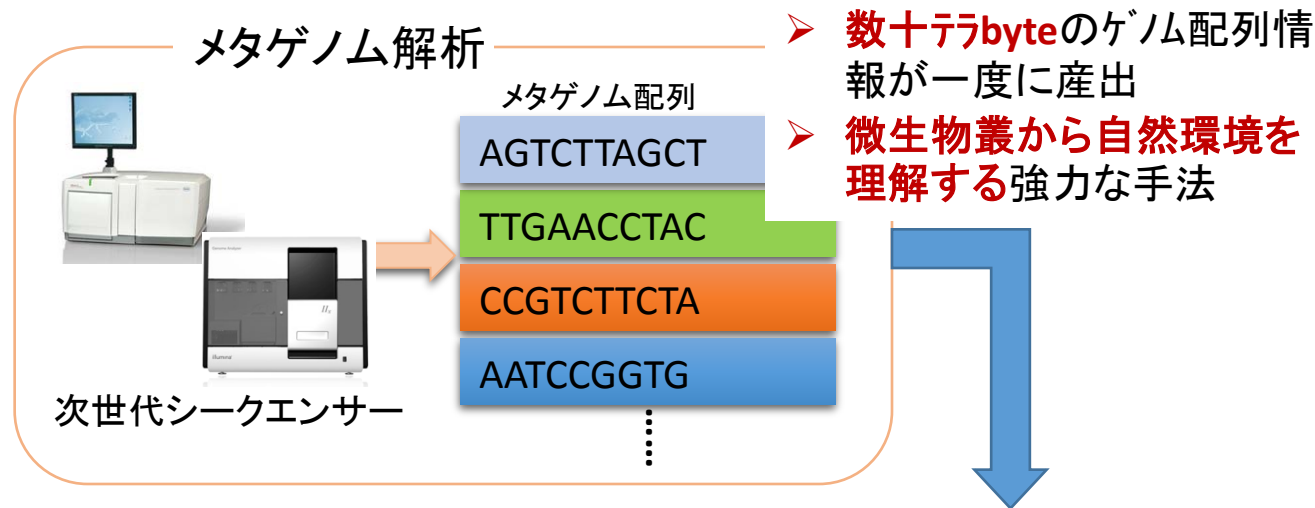
<計画的に利用できていない場合、その理由>

より高速化を目指した大規模ゲノム解析のための自己圧縮 BLSOM のアルゴリズム、ならびに、条件検討に時間がかかり、自己圧縮 BLSOM の大規模解析を行うことができなかった。

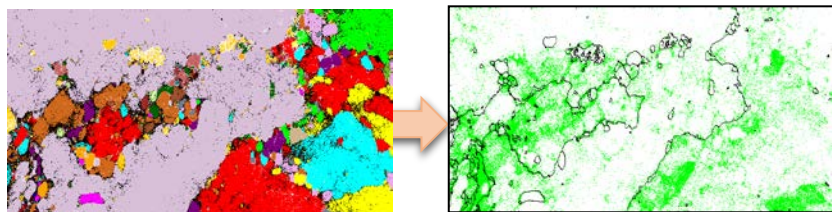
8. 新聞、雑誌での掲載記事

特になし。

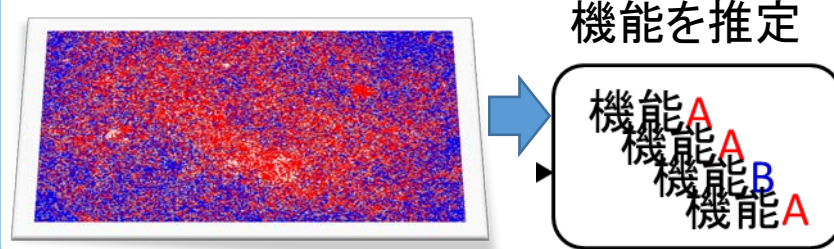
BLSOM(一括学習型自己組織化マップ)を用いたメタゲノム配列解析



BLSOMによる生物系統推定



BLSOMによる遺伝子機能推定



未知ゲノム(メタゲノム配列)へ自然環境を把握するための有用な情報を付加するには、**全既知情報を元にしたリファレンスの作成**が重要

ゲノム配列データは加速度的に増加

高速化アルゴリズムが完成し、より大量なゲノム配列データを対象にした解析が可能